Efficient Extraction of Human Motion Volumes by Tracking

Juan Carlos Niebles\textsuperscript{1,2} \quad Bohyung Han\textsuperscript{3} \quad Li Fei-Fei\textsuperscript{4}

\textsuperscript{1}Princeton University, USA  \quad \textsuperscript{2}Universidad del Norte, Colombia

\textsuperscript{3}UNIST, Korea \quad \textsuperscript{4}Stanford University, USA

\texttt{jniebles@princeton.edu} \quad \texttt{bhhan@unist.ac.kr} \quad \texttt{feifeili@cs.stanford.edu}

**Summary**

Goal: Efficiently extract the spatio-temporal volume that encloses each person in a video.

Contributions:
- A system designed to automatically and quickly extract human motion volumes from challenging videos by combining top-down and bottom-up methods.
- A top-down technique to obtain a probabilistic human body contour.
- A global optimization procedure based on belief propagation for bottom-up processing.

**People Detection & Clustering**

- Collect hypotheses from a pedestrian detector.
- Impose spatio-temporal location and appearance constraints to cluster detections. \cite{Klein ICML '02}
- Each cluster contains detections from a unique individual.

**Experimental Results**

- 50 \texttt{YouTube} videos from \cite{Niebles CVPR 2010}
- Bottom-up processing \textasciitilde50ms.
- Top-down estimation \textasciitilde1s.

**Discussions**

- System relies on observing at least a few frames where the subject is in a upright position.
- Fast moving limbs are still the most difficult to track.
- Multiple-target tracking to better deal with occluding subjects.